MASRI — Corpora abertos de reconhecimento de fala da Universidade de Malta para uma língua com cerca de 500 000 falantes, disponíveis só para uso não comercial

Digital Inclusion · Boa prática · Malta

Pontuação de qualidade: 72/100

Força da evidência: Forte

Sumário

O MASRI criou os primeiros corpora de reconhecimento de fala em maltês para que as interfaces de voz funcionem em maltês e não só em inglês: um corpus de 8 horas com revisão por pares (25 falantes) cresceu para cerca de 142 horas. A licença não comercial limita o uso em produtos.

Dimensões-chave

Dimensão	Pontuação
Innovation level	3/5
Sustainability (ongoing)	1/1
Evaluation evidence	1/1
Demonstrated reach	0/1
Documented learning	1/1

Avaliado por C-NAPSE

Fontes de dados

C-NAPSE web research — acedido a 2026-11-19

University of Malta — Institute of Linguistics and Language Technology — acedido a 2026-11-19

University of Malta — MASRI project page — acedido a 2026-11-19

Mena et al. (2020), MASRI-HEADSET: A Maltese Corpus for Speech Recognition, LREC 2020 (arXiv) — acedido a 2026-11-19

MASRI-Dev evaluation set on Hugging Face (MLRS) — acedido a 2026-11-19

Citar — Evidoria. *MASRI — Corpora abertos de reconhecimento de fala da Universidade de Malta para uma língua com cerca de 500 000 falantes, disponíveis só para uso não comercial*. ID persistente: 20dcd978-f906-46c6-9479-e19936878200.